MATTEO RIONDATO
Curriculum vitae

C214 Science Center
25 East Drive
Amherst, MA 01002, USA
rionda@acm.org
http://matteo.rionda.to

EDUCATION

Ph.D.	Computer Science, Brown University, 2014
M.Sc.	Computer Science, Brown University, 2010
M.Sc. (Laurea Specialistica)	Computer Engineering, University of Padua (Italy), 2009
B.Sc. (Laurea)	Information Engineering, University of Padua (Italy), 2007

ACADEMIC APPOINTMENTS

2019–	Assistant Professor, Department of Computer Science, Amherst College
2016–	Visiting Faculty (various titles), Department of Computer Science, Brown University
2015	Postdoctoral Research Associate, Department of Computer Science, Brown University
2014–15	Postdoctoral Researcher, Department of Computer Science, Stanford University

INDUSTRY POSITIONS

2015–18	Research Scientist, Labs, Two Sigma Investments LP
2013	Summer Intern Research Scientist, Web Mining Group, Yahoo Research Barcelona

PUBLICATIONS
Authors in alphabetical order unless marked otherwise (*)

Journal Articles

(to appear)	**M. Riondato** and F. Vandin. MiSoSouP: Mining Interesting Subgroups with Sampling and Pseudodimension. *ACM Transactions on Knowledge Discovery from Data,* **Invited article to the special issue of the best papers from KDD**

2018

(to appear)   * S. Servan-Schreiber, **M. Riondato**, and E. Zgraggen. ProSecCo: Progressive Sequence Mining with Convergence Guarantees, *Knowledge and Information Systems*, **Invited article to the special issue on the best papers from ICDM 2018**

2019   C. Cousins and **M. Riondato**. CaDET: Interpretable Parametric Conditional Density Estimation with Decision Trees and Forests. *Machine Learning*, 108:1613–1634

2018   **M. Riondato** and E. Upfal. ABRA: Approximating Betweenness Centrality in Static and Dynamic Graphs with Rademacher Averages. *ACM Transactions on Knowledge Discovery from Data*, 12(5):61:1–61:38

2017   L. De Stefani, A. Epasto, **M. Riondato**, and E. Upfal. TRIÉST: Counting Local and Global Triangles in Fully-dynamic Streams with Fixed Memory Size. *ACM Transactions on Knowledge Discovery from Data*, 11(4):43:1–43:50, **Invited article to the special issue on the best papers from KDD 2016**

2017   * **M. Riondato**, D. García-Soriano, and F. Bonchi. Graph Summarization with Quality Guarantees. *Data Mining and Knowledge Discovery*, 31(2):314–349

2015   * **M. Riondato** and E. M. Kornaropoulos. Fast Approximation of Betweenness Centrality through Sampling. *Data Mining and Knowledge Discovery*, 30(2):438–475

2014   **M. Riondato** and E. Upfal. Efficient Discovery of Association Rules and Frequent Itemsets through Sampling with Tight Performance Guarantees. *ACM Transactions on Knowledge Discovery from Data*, 8(4):20:1–20:32

2010   A. Pietracaprina, **M. Riondato**, E. Upfal, and F. Vandin. Mining Top-k Frequent Itemsets through Progressive Sampling. *Data Mining and Knowledge Discovery,* 21(2):310–326, I**nvited article to the special issue on the best papers from ECML PKDD 2010**

Conference Proceedings

2020   * L. Pellegrina, C. Cousins, F. Vandin, and **M. Riondato**. McRapper: Monte-Carlo Rademacher Averages for Poset Families and Approximate Pattern Mining. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), to appear.

2019   L. Pellegrina, **M. Riondato**, and F. Vandin. SPuManTE: Significant Pattern Mining with Unconditional Testing. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 1528-1538

2019   L. Pellegrina, **M. Riondato**, and F. Vandin. Hypothesis testing and statistically-sound pattern mining. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 3215-3216

2018   * S. Servan-Schreiber, **M. Riondato**, and E. Zgraggen. ProSecCo: Progressive

Sequence Mining with Convergence Guarantees. *Proceedings of the 18th IEEE International Conference on Data Mining* (ICDM), 417–426, **Best Student Paper Award runner-up**

2018	**M. Riondato** and F. Vandin. MiSoSouP: Mining Interesting Subgroups with Sampling and Pseudodimension. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 2130–2139
2016	L. De Stefani, A. Epasto, **M. Riondato**, and E. Upfal. TRIÉST: Counting Local and Global Triangles in Fully-dynamic Streams with Fixed Memory Size. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 825–834, **Best Student Paper Award (Research Track)**
2016	**M. Riondato** and E. Upfal. ABRA: Approximating Betweenness Centrality in Static and Dynamic Graphs with Rademacher Averages. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 1145–1154
2016	F. Bonchi, G. De Francisci Morales, **M. Riondato**, Centrality Measures on Big Graphs: Exact, Approximated, and Distributed Algorithms. *Proceedings of the 25th International World Wide Web Conference* (WWW), Companion Volume, pp. 1017–1020
2016	A. Mahmoody, **M. Riondato**, and E. Upfal. Wiggins: Detecting Valuable Information in Dynamic Networks with Limited Resources. *Proceedings of the 9th ACM International Conference on Web Search and Data Mining* (WSDM), pp. 677–686
2015	**M. Riondato** and E. Upfal. Mining Frequent Itemsets through Progressive Sampling with Rademacher Averages. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 1005–1014
2015	**M. Riondato** and E. Upfal. VC-Dimension and Rademacher Averages: From Statistical Learning Theory to Sampling Algorithms. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD), pp. 2321–2322
2015	A. Anagnastopoulos, L. Becchetti, A. Fazzone, I. Mele, and **M. Riondato**. The Importance of Being Experts: Efficient Max-Finding in Crowdsourcing. *Proceedings of the 36th ACM SIGMOD International Conference on Management of Data* (SIGMOD), pp. 983–998
2014	* **M. Riondato**, D. García-Soriano, and F. Bonchi. Graph Summarization with Quality Guarantees. *Proceedings of the 14th IEEE International Conference on Data Mining* (ICDM), pp. 947–952
2014	**M. Riondato**. Sampling-based Data Mining Algorithms: Modern Techniques and Case Studies. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML PKDD),

pp. 516–519

2014	**M. Riondato** and F. Vandin. Finding the True Frequent Itemsets. *Proceedings of the 14th SIAM International Conference on Data Mining* (SDM), pp. 497–505
2014	*** M. Riondato** and E. M. Kornaropoulos. Fast Approximation of Betweenness Centrality through Sampling. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* (WSDM), pp. 413–422
2012	*** M. Riondato**, J. A. DeBrabant, R. Fonseca, and E. Upfal. PARMA: A Parallel Randomized Algorithm for Association Rules Mining in MapReduce. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (CIKM), pp. 85–94
2012	**M. Riondato** and E. Upfal. Efficient Discovery of Association Rules and Frequent Itemsets through Sampling with Tight Performance Guarantees. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML PKDD), pp. 25–41
2012	A. Pietracaprina, G. Pucci, **M. Riondato**, F. Silvestri, and E. Upfal. Space-round Tradeoffs for MapReduce Computations. *Proceedings of the 26th ACM International Conference on Supercomputing* (ICS), pp. 235–244
2012	M. Akdere, U. Çetintemel, **M. Riondato**, E. Upfal, and S. B. Zdonik. Learning-based Query Performance Modeling and Prediction. *Proceedings of the 28th IEEE International Conference on Data Engineering* (ICDE), pp. 390–401
2011	*** M. Riondato**, M. Akdere, U. Çetintemel, S. B. Zdonik, and E. Upfal. The VC-dimension of SQL Queries and Selectivity Estimation through Sampling. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML PKDD), pp. 661–676
2011	M. Akdere, U. Çetintemel, **M. Riondato**, E. Upfal, and S. B. Zdonik. The Case for Predictive Database Systems: Opportunities and Challenges. *Proceedings of the 5th Biennial Conference on Innovative Data System Research* (CIDR), pp. 167–174

Technical Reports

2018	**M. Riondato**. Sharpe Ratio: Estimation, Confidence Intervals, and Hypothesis Testing. *Two Sigma Technical Report Series*, No. 2018-001.

Tutorials

2020	**M. Riondato**. Hypothesis testing and statistically-sound pattern mining. SIAM SDM'20 (co-organized with L. Pellegrina and F. Vandin) *(Conference canceled due to COVID-19)*
2019	L. Pellegrina, **M. Riondato,** F. Vandin. Hypothesis testing and statistically-sound pattern mining. ACM KDD'19
2016	F. Bonchi, G. De Francisci Morales, **M. Riondato**. Centrality Measures in Big

Graphs: Exact, Approximated, and Distributed Algorithms. WWW'16

2015   **M. Riondato** and E. Upfal. VC-Dimension and Rademacher Averages: From Statistical Learning Theory to Sampling Algorithms. ACM KDD'15, ECML PKDD'15, ACM CIKM'15

Abstracts

2020   **M. Riondato.** Algorithms for Scalable Hypothesis Testing. SIAM Conference on Mathematics of Data Science *(Conference postponed due to COVID-19)*

AWARDS

2018   Invited paper to the special issue of Knowledge and Information Systems for the best papers of IEEE ICDM'18

2018   Best Student Paper Award runner-up at IEEE International Conference on Data Mining (IEEE ICDM'18)

2018   Invited paper to the special issue of ACM Transactions on Knowledge Discovery from Data for the best papers of ACM KDD'18

2018   Invited Young Researcher to the 6th Heidelberg Laureate Forum

2016   Best Student Paper Award (Research Track) at ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM KDD'16)

2014   Nominated for the ACM SIGKDD Doctoral Dissertation Award

2014   Best Student Poster Award at SIAM International Conference on Data Mining (SDM'14)

2010   Paper selected for the special issue of Data Mining and Knowledge Discovery for the best papers of ECML PKDD'10

GRANTS AND FELLOWSHIPS

2019   PI on NSF Award #1918446: NSF Student Travel Grant for 2019 SIAM International Conference on Data Mining (SDM) (USD 25k)

2015   SIAM/NSF Early Career Travel Award to SIAM Intl. Conference on Data Mining

2014   SIAM Travel Award to SIAM Intl. Conference on Data Mining

2014   Brown University Dissertation Fellowship

2013   Yahoo Research Barcelona Summer Internship

2011   Italy MIUR Research Fellowship

2009   Brown University Graduate Fellowship

INVITED TALKS

2020   Algorithms for centrality measures, CS591 – Large-Scale Graph Mining, Boston University, Boston (MA, USA), April 13

2019   SPuManTE: Significant Pattern Mining with Unconditional Testing, Network Science Institute, Northeastern University, Boston (MA, USA), November 8

2019   Making better use of data, MassMutual Research Bytes series, Amherst (MA, USA), November 5

2019   Making better use of data, Data Management Lab, Boston University, Boston (MA, USA), October 25

2019   Making better use of data, Department of Computer Science, Mount Holyoke College, South Hadley (MA, USA), October 16

2019   CaDET: Interpretable Parametric Conditional Density Estimation with Decision Trees, Workshop on Computation and Statistics in Data Science (CaStleD'19), Bertinoro (FC, Italy), October 1

2019   Pseudodimension for Data Analysis, Workshop on Data Science in Low-dimensional Spaces, ICERM, Brown University, Providence (RI, USA), May 17

2018   MiSoSouP: Mining Interesting Subgroups with Sampling and Pseudodimension, Google Research NY, New York (NY, USA), December 4

2018   Approximation Algorithms for Betweenness Centrality, COMPSCI 134 – Networks, Harvard University, Cambridge (MA, USA), November 26

2018   Data Mining: Tasks, Systems, Challenges, and Research Directions, Amherst College, Amherst (MA, USA), November 5

2018   Statistical Learning Theory meets Data Mining: Fast, High-quality, Sampling-based Approximation Algorithm, ISI Foundation, Turin (Italy), September 8

2018   Sampling-based Approximation Algorithms for Data Analysis using Rademacher Averages, Theory of Computation Seminar, Harvard University, Cambridge (MA, USA), May 7

2018   Statistical Learning Theory meets Data Mining: Fast, High-quality, Sampling-based Approximation Algorithms, Boston College, Chestnut Hill (MA, USA), January 31

2018   Statistical Learning Theory meets Data Mining: Fast, High-quality, Sampling-based Approximation Algorithms, Amherst College, Amherst (MA, USA), January 25

2017   Betweenness Centrality Estimation with Rademacher Averages, National Institute of Informatics, Tokyo (Japan), November 16

2017   Betweenness Centrality Estimation with Rademacher Averages, Center for Data Science, New York University, New York (NY, USA), May 17

2017   Rademacher Averages: Theory and Practice, Dagstuhl Seminar 17141, Schloss Dagstuhl

(Germany), April 6

2017   Random Sampling for Data Mining: The Case of Triangles in Dynamic Streams, School of Computer Science, McGill University, Montreal (QB, Canada), February 21

2017   Random Sampling for Data Mining: The Case of Triangles in Dynamic Streams, College of Computer and Information Science, Northeastern University, Boston (MA, USA), February 15

2017   The Neverending Data – Streaming, Sampling, and Triangle Counting, Department of Computer Science, Amherst College, Amherst (MA, USA), January 27

2016   Approximating Betweenness Centrality through Sampling with the Rademacher Averages, Data Management Lab, Boston University, Boston (MA, USA), November 18

2016   Algorithmic Data Science = Theory + Practice, IEEE MIT Undergraduate Technology Research Conference, Massachusetts Institute of Technology, Cambridge (MA, USA), November 5

2016   TRIÉST: Counting Local and Global Triangles in Fully-dynamic Streams with Fixed Memory Size, Database Group, Carnegie Mellon University, Pittsburgh (PA, USA), October 24

2016   Approximating Betweenness Centrality through Sampling with the Rademacher Averages, Network Science Institute, Northeastern University, Boston (MA, USA), October 17

2016   Graph Summarization with Quality Guarantees, Department of Information Engineering, University of Padua, Padua (Italy), September 26

2016   ABRA: Venice, Sampling, and Betweenness Centrality Estimation, Social Impact through Network Science (SINS), Venice (Italy), June 8

2015   Data is …, Stevens Institute of Technology, Hoboken (NJ, USA), December 12

2015   Travel Pictures from Another World: Statistical Learning Theory Meets Data Mining, Monash University, Melbourne (Australia), October 28

2015   Modern Sampling for Modern Data: The Case of Frequent Itemsets Mining, Two Sigma Investments, New York (NY, USA), March 5

2014   Statistical Learning and Data Mining: A Lasting Marriage, Data Mining: Beyond The Horizon Workshop, University of Bristol, Bristol (UK), November 20

2014   Efficient Frequent Itemsets Mining through Sampling, Database Research Group, University of Waterloo, Waterloo (ON, Canada), May 7

2014   Efficient Frequent Itemsets Mining through Sampling, Database Group, MIT CSAIL, Cambridge (MA, USA), April 10

2014   Taming the Challenges of Big Data with Statistical Data Analytics, Department of Computer Science, Boston College, Boston (MA, USA), February 7

2013   Fast Betweenness Estimation through Sampling, Data Management Lab, Boston University, Boston (MA, USA), October 17

2013   Fast Betweenness Estimation through Sampling, Yahoo! Labs Barcelona, Barcelona

(Spain), June 13

2013   Fast Betweenness Estimation through Sampling, Advanced Computing Group, University of Padua, Padua (Italy), May 30

2013   Statistical Learning Theory meets Knowledge Discovery, Brown CS Industrial Partners Program Symposium, Providence (RI, USA), April 25

2011   Statistical Learning Theory meets Databases, Advanced Computing Group, University of Padua, Padua (Italy), June 24

2010   Top-k Frequent Itemsets Mining through Sampling, Advanced Computing Group, University of Padua, Padua (Italy), September 16

2010   Mining Top-K Frequent Itemsets Through Progressive Sampling, Department of Computer Science, Chalmers University of Technology, Gothenburg (Sweden), August 5

ADDITIONAL RESEARCH EXPERIENCE AND VISITS

2017   National Institute of Informatics (Tokyo, Japan) – Visiting Researcher, November

2015   Monash University (Melbourne, Australia) – Visiting Researcher, October

2012   Sapienza University of Rome (Rome, Italy) – Visiting Ph.D. Student, June–September

2011   University of Padua (Padua, Italy) – Research Fellow, June–September

2010   Chalmers University of Technology (Gothenburg, Sweden) – Visiting Ph.D. Student, August

2008–9   Brown University (Providence, RI, USA) – Visiting Student, October–June

TEACHING EXPERIENCE

Amherst College

Advanced Analytics COSC-490 (independent study, 1 student, spring 2020)

Databases COSC-257 (fall 2019)

Database Implementation COSC-490 (independent study, 2 students, spring 2020)

Data Mining COSC-254 (spring 2019)

Introduction to Computer Science I COSC-111 (spring 2019, fall 2019, spring 2020)

Network Science COSC-355 (spring 2020)

COSC-490 Regression and Time Series (independent study, 1 student, fall 2019)

Theory meets Practice STATS-490H (fall 2019)

Brown University

Optimization Methods in Finance CSCI-1951-G (spring 2016, spring 2018)

Probability and Computing CSCI-1550 (Teaching Assistant) (fall 2010, spring 2012, spring 2013, spring 2014)

STUDENT SUPERVISION

Cyrus Cousins, research intern, Labs team, Two Sigma Investments, Summer 2018

Conrad Kuklinsky, research assistant, Amherst College, Spring 2019 – Ongoing

Kathleen Isenegger, honors thesis, Amherst College, 2019-20

Zhiyuan Jia, research assistant, Amherst College, Summer 2019

Sacha Servan-Schreiber, research assistant, Brown University, Spring – Fall 2018

Shukry Zablah, honors thesis, Amherst College, 2019-20

SERVICE TO THE SCIENTIFIC COMMUNITY

Editorial Boards

Data Mining and Knowledge Discovery

Machine Learning (Guest editorial board for ECML PKDD '20, '19)

Organizing Committees

SIAM SDM '21 (Tutorial Chair)

SIAM SDM '20, '19 (Doctoral Forum Co-chair)

SIAM SDM '18 (Sponsorship Co-chair)

Foundations of Learning from Data Workshop (Bertinoro, Italy, September 2018)(Co-Organizer)

Program Committees

AAAI '20, '19

ACM CIKM '20, '19, '18, '17, '16, '15, '14

ACM KDD '20, '19, '18, '17 (Senior PC), '16, '15

ACM WSDM '20, '19, '17

ECML PKDD '20, '19, '18, '17, '16

Grace Hopper '18, '16

ICML '19

IEEE ICDE '18, '17

IEEE ICDM '20, '19, '18 '16

IJCAI '19

NetSci '20, '19

NeurIPS '20, '19

SIAM SDM '20, '19

WWW '20, '19, '18, '17, '16

Journal Reviewing

ACM Transactions on Database Systems (TODS)

ACM Transactions on Information Systems (TOIS)

ACM Transactions on Knowledge Discovery from Data (TKDD)

Data Mining and Knowledge Discovery (DMKD/DAMI)

Discrete Applied Mathematics (DAM)

Engineering Applications of Artificial Intelligence (EAAI)

Expert Systems with Applications (ESWA)

Journal of Parallel and Distributed Computing (JPDC)

Knowledge and Information Systems (KAIS)

IEEE Access

IEEE Open Access Journal of the Computer Society (OJCS)

IEEE Transactions on Knowledge and Data Engineering (TKDE)

IEEE Transactions on Network Science and Engineering (TNSE)

IEEE Transactions on Parallel and Distributed Systems (TPDS)

IEEE Transactions on Services Computing (TSC)

Information Science (INS)

Machine Learning (Mach. Learn.)

PeerJ Computer Science

PLOS ONE

The Computer Journal

VLDB Journal (VLDBJ)

The World Wide Web Journal (WWWJ)

Conference Reviewing
The list does not include reviews done as Program Committee member

LATIN'20, MFCS'17, SODA'17, ISAAC'15, SIAM SDM'15, IEEE ICDE'15, DISC'14, ACM WSDM'14, WWW'14, ICALP'14, IEEE BigData'13, MFCS'13, ACM WSDM'13, IEEE IPDPS'12, ACM ICS'12, RANDOM'11

Conference Research Session Chair

WWW'20 – Social Network Analysis III

SDM '19 – Patterns

ACM KDD '18 – Unsupervised Learning II

ACM KDD '17 – Graphs I

ACM KDD '16 – Graphs I

IEEE ICDM '16 – Theory

ECML PKDD '16 – Graphs and Social Networks 1

WWW '16 – Social Networks and Graph Analysis 1

ACM KDD '15 – Social and Graphs 4

Grant Reviewing

NSF SBIR/STTR Program

Sigma Delta Epsilon (Graduate Women in Science) National Fellowship

Grace Hopper Conference Scholarship 2016, 2017

DEPARTMENTAL / UNIVERSITY SERVICE

Brown University Graduate Council – Student Representative, 9/2011 – 9/2013

Brown University Graduate Student Council – President, 4/2011 – 12/2012

Brown University Strategic Planning Committee on Doctoral Education – Student Representative, 9/2012 – 5/2013

Brown University New Scientist Program Graduate–Undergraduate Mentoring Initiative – Graduate Mentor, 9/2013 – 5/2014

ADDITIONAL TRAINING

2014   Brown University Harriet W. Sheridan Center for Teaching and Learning
        Teaching Certificate I: Reflective Teaching

2012   Summer School on Massive Data Mining, IT University, Copenhagen, Denmark

PROFESSIONAL MEMBERSHIPS

Association for Computing Machinery (ACM), 2012 – present
        Special Interest Group on Knowledge Discovery from Data (SIGKDD)
        Special Interest Group on Management of Data (SIGMOD)
        Special Interest Group on Algorithms and Computation Theory (SIGACT)

Institute of Electrical and Electronic Engineers (IEEE), 2012 – present
        IEEE Computer Society

Society for Industrial and Applied Mathematics (SIAM), 2012 – present
        SIAM Activity Group on Data Mining and Analytics (SIAG/DMA)

The latest revision of this CV is available from http://matteo.rionda.to/MatteoRiondato-CV.pdf
This revision was created on May 18, 2020